

Preface

December 18, 2000, the date the IBM eServer z900 was introduced, marked a significant milestone in the evolution of IBM mainframes. Although it was built on the legacy of its CMOS-chip-based predecessors—the S/390® Parallel Enterprise Servers G3 through G6—the z900 was very different. It was the first of a series of mainframe servers developed specifically to meet the unique computing challenges presented by e-business.

Every major subsystem was rethought, redefined, and redesigned—all with e-business requirements in mind. A new system architecture, z/Architecture™, was introduced to significantly improve mainframe capacity. Although upward compatible with programming written for prior S/390 mainframe architectures, z/Architecture introduces full 64-bit capability—real and virtual addressing and arithmetic.

The microprocessor of the z900 was redesigned not only to provide support for the new 64-bit architecture, but also to build upon learning from past designs so as to provide additional performance and function enhancements. The unique binodal cache system structure was enhanced to enable attachment of up to twenty microprocessors, significantly extending total system performance.

To ensure well-balanced operation, the I/O subsystem was completely redesigned. The I/O bandwidth was tripled compared to that of the G6, and packaging densities of high-performance ESCON® and FICON™ channels were respectively quadrupled and doubled. HiperLink bandwidth was doubled, while Integrated Cluster Bus (ICB) bandwidth was tripled. A new type of communications interface, HiperSockets, was introduced to provide very-high-performance communications between logical partitions (LPARs). The end result of these design changes was a completely redesigned e-business server, introduced just 18 months after its predecessor, yet offering more than 50 percent more in total computing capacity.

However, e-business is not just about performance and bandwidth. Success in e-business is defined by continuous operation and availability. The z900 is a self-protecting, self-configuring, self-optimizing, self-healing server. At the time of the introduction of the z900, it was the most advanced embodiment yet of the IBM self-management technologies. Among these are the Intelligent Resource Director, which provides the z900 the capability of moving required computational and I/O resources where they are needed most, and improved concurrent capacity upgrade and downgrade functions. Fault detection and recovery capabilities have been enhanced over those of predecessor systems. The service and system controls infrastructure was redesigned, providing improved performance and availability. The list goes on. This combination of high

performance and autonomic capabilities is directly responsible for the significant revenue growth the mainframe platform has experienced.

This double issue of the *IBM Journal of Research and Development* contains 19 papers that describe many aspects of the z900, ranging from its architecture and design to its leading-edge implementation of several autonomic computing concepts. Thanks are due to the many authors from the Global Hardware and Software Laboratories of the IBM Server Group, the IBM Microelectronics Division, and the IBM Thomas J. Watson Research Center who have taken the time to document these outstanding accomplishments.

The paper by Plambeck et al. describes the development of the new z/Architecture. The single most important new capability introduced in the z900, the challenge during z/Architecture development, was to enable the full range of 64-bit real and virtual addressing and 64-bit arithmetic while also providing full upward compatibility for earlier versions of operating systems, middleware, and applications. The result is an interesting and unique architecture which permits 64- and 31-bit addressing to be intermixed with 64- and 32-bit arithmetic modes.

The implementation of z/Architecture in the microprocessor core is discussed in the paper by Schwarz et al. The paper also addresses other important performance enhancements, including a redesign of the L1 cache, a new translator design, and a new hardware compression unit which improves compression performance by a factor of 3 and expansion performance by a factor of 2.

The desire to extend the unique binodal cache system structure, in order to increase overall Symmetrical Multi-Processor (SMP) size, led to significant wiring and frequency challenges to the electronic packaging used. The result is the densest electronic package in the industry, comprising 35 chips and 1 km of wiring, supporting chip-to-chip frequencies of up to 459 MHz, all on a package 127 mm on a side. Harrer et al. describe not one, but two, designs which successfully address this complex design point.

Total computing capacity in excess of 2600 MIPS demands significant I/O bandwidth. Consequently, the I/O subsystem was completely redesigned. Most significant was the introduction of the 1GB/s Self-Timed Interface (STI) as the backbone of the I/O subsystem. Stigliani et al. present an overview of the entire I/O subsystem, including new ESCON, FICON, and networking adapters, while Hoke et al. provide a more detailed description of the design and operation of the STI circuitry itself.

The new STI design also provided an opportunity to dramatically upgrade the overall performance of the zSeries Parallel Sysplex® functions. The InterSystem Channel (ISC) and Integrated Cluster Bus (ICB) were both redesigned to take full advantage of the greater bandwidth afforded by the new STI. The paper by Gregg

and Errickson describes these improvements in more detail.

The z900 also fully exploits a messaging architecture, Queued Direct I/O (QDIO), first introduced on G5/G6-class servers. The architecture forms the basis for the new high-performance Ethernet, Token Ring, and ATM adapters as well as a new function designated as HiperSockets, developed to improve communications efficiency between logical partitions. HiperSockets, because it is implemented totally in Licensed Internal Code (LIC), provides a high-performance capability which is expected to be an important enabler to the Linux[®] initiatives of the z900. Another important Linux enablement is the support for the Fibre Channel Protocol (FCP) attachment to the z900, a major departure from the traditional mainframe Extended Count Key Data (ECKD) protocol. The paper by Baskey et al. describes the HiperSockets messaging architecture, while the paper by Adlung et al. describes the FCP implementation in the z900.

The next group of papers discusses several aspects of the implementation of self-managing technologies in the z900. First, the z900 is built upon a strong RAS legacy. The paper by Alves et al. provides a survey of the key RAS advances and enhancements, focusing on the design objective of continuous reliable operation and self-managing computing. The service and system controls infrastructure is a critical, but often overlooked, subsystem. The paper by Baitinger et al. describes the service subsystem topology of the z900, including a description of the hardware abstraction required for software control. The paper by Bieswanger et al. expands upon the hardware object model, describing its operation as well as application and user interface design. Probst et al. address the implementation in the z900 of new self-configuring capabilities, while the paper by Valentine et al. describes the high-availability aspects of the design of the z900 Service Console.

Rooney et al. provide a description of an innovative approach to workload management—using customer-defined priorities and rules to manage computing and I/O resources across the system; in effect, allocating resources on the fly to the workloads that need them most. Notably, this combination of new and enhanced functions provides the z900 with unmatched capabilities, a harbinger of true autonomic computing.

Designing a server of such complexity would not have been possible without the use of innovative techniques in design verification and high-frequency design. Verification capability is critical to achieving time-to-market objectives with superior quality, particularly when considering the massive amount of new function introduced in the z900. Koerner et al. present an overview of the Virtual Power-On process used to validate system microcode.

This methodology was credited with saving weeks of development effort, contributing to early system release. The paper by Kayser et al. provides more detail on the simulation accelerator, and the challenge of verifying hardware and software within the same simulation environment.

Von Buttler et al. address the challenge of software and firmware simulation. They describe CECSIM—a significant advance in simulation capability which uses some of the unique capabilities of the VM operating system to create a very powerful, very flexible simulation environment. The paper on functional verification by Silverio et al. describes in detail the simulation environment created to verify a critical I/O subsystem component in record time with excellent quality.

The final paper, by Curran et al., presents an overview of the design methodology, circuit design techniques, and chip layout and design approach used to produce a complex microprocessor capable of executing at frequencies in excess of 1 GHz.

Paul R. Turgeon
Program Manager, eServer Hardware Development
IBM Server Group

Guest Editor